

# A Big Data Analytics Framework in Climate Smart Agriculture

Tekalign Tujo Gurmesssa

Lecturer, Department of Computer Science, College of Computing, Madda Walabu University

PO. Box 247, Bale Robe, Ethiopia

## Abstract

Climate Smart Agriculture incorporates information on soils, bothers, maladies, costs and different variables to increment illustrative power. Creating atmosphere strong horticulture is fundamental to accomplishing future sustenance security and environmental change objectives. Through CSA application the ranchers can anticipate crop type (crop exhortation) which is suitable for accessible condition. Climate Smart farming has been generally used to express agricultural practices that will increment horticultural efficiency and nourishment security and to foresee rural items. CSA application have Climate information (like: Max Temp, Min Temp, Humidity, Rain fall, daylight, wind course, wind speed), Fertilizer and Soil Data's (like, Black Soil, Red Soil).The review aim to help farmers better adapt to temperature extremes, droughts or excess water in fields so that they can make better decisions for the environment and maximize production or profits. The data collection is an important role in the work process. Enabling farmers to head massive amounts of data collected through sensors to predict the best time to plant, what type of seed to use, and where to plant in order to improve yields, cut operational costs, and minimize environmental impact. Big data analytics provide new ways for businesses updates and requirement for updating and government to analyze unstructured data. Now a day, big data is one of the most important and challenging point in information technology world. It is executing very important role in future. Big data changes the way of world for management and use big amount of data

**Keywords:** Climate Smart Agriculture, Big Data Analytics and Hadoop

**DOI:** 10.7176/CEIS/10-6-01

**Publication date:** July 31<sup>st</sup> 2019

## Introduction

The use of technology in agriculture has continued to grow since the early part of the 20<sup>th</sup> century, when the industry shifted from the horse-drawn plow to mechanized tractors. The advent of plant genetics, chemical inputs, and, more recently, guidance systems has transformed the industry into one that is increasingly technology-intense and data-rich. The ability to generate, capture, and store data in the agricultural industry has continued to grow with the use of mobile technology and data management software. Additionally, external data sets are now readily available to the industry, allowing for a more complete picture of the world in which production agriculture occurs. The technological advances that make up the modern computing environment have contributed to debate about big data. While data collecting is not new, especially in the context of public data collection, only since the advent of more efficient, mobile technologies and the digitization of data have large records been able to be evaluated and analyzed in a timely and more useful way. Climate-smart agriculture includes proven practical techniques and approaches that can help achieve a triple win for food security, adaptation and mitigation[1] (FOA, 2010). Climate-smart agriculture fully incorporates attention to climate risk management. CSA incorporates data on soils, pests, diseases, costs and other factors to increase explanatory power.

According to [2] (Sara J Scherr\*, 2012), Climate-smart agricultural landscapes is the integrated landscape approach offers a strategy to achieve climate-smart agriculture objectives at scale and in all its dimensions. Through climate-smart agricultural landscapes, important synergies for agricultural production, climate adaptation and mitigation, as well as other livelihood and environmental objectives, can be generated through coordinated action at farm and landscape scales.

This review is intended to address research challenges in Agriculture sector, a new and interdisciplinary research area that spans the Big Data Analytics, Data Science tools, Hadoop, R and Machine learning algorithms. To help agriculture sectors in produce more yields. Big data holds one of the keys for farmers control centers collect and process data in real time to help farmers make the best decisions with regard to planting, fertilizing and harvesting crops.

## Big Data Analytics

Big Data Analytics is aimed at making sense of data by applying efficient and scalable algorithms on Big Data for its analysis, learning, modelling, visualization and understanding. This includes the design of efficient and effective algorithms and systems to integrate the data and uncover the hidden values from data. It also includes methodologies and algorithms for automatic or mixed-initiative knowledge discovery and learning, data transformation and modelling, predictions and explanations of the data. Breakthroughs in this area include new algorithms, methodologies, systems and applications for knowledge discovery, understanding and applications

based on the Big Data. New computing paradigms are expected in new areas such as human computation, crowd sourcing, and sentiment analysis as well as data visualization technologies.

[3] (Chan, 2013) Provided a classification of business intelligence and analytics (BI&A) into three categories. BI&A 1.0 is characterized by DBMS-based and structured content. It utilizes traditional analytic tools via data warehousing, ETL, OLAP and data mining. BI&A 2.0 is characterized by Web-based and unstructured content. It utilizes tools in information retrieval, opinion mining, question answering, Web analytics, social media analytics, social network analysis, and spatial-temporal analysis. BI&A 3.0 is characterized by mobile and sensor-based content. It utilizes tools in location-awareness analysis, person-centered analysis, context-relevant analysis, and mobile visualization and HCI. BI&A 2.0 and 3.0 would require a platform that can handle the huge volume, velocity and variety of data. The Big Data analytics architecture described below utilizes the massively parallel, distributed storage and processing framework as provided by Hadoop HDFS and MapReduce.

As opposed to some belief that Big Data has pronounced the obsolescence of data warehousing, it remains a viable technology for Big Data analytics of huge volume of structured data. Furthermore, there is synergy between data warehousing and the Hadoop type Big Data architecture. Map Reduced data can then be integrated with the data warehouse for further analytic processing. Conversely, data warehouse can be a data source for complex Hadoop jobs, simultaneously leveraging the massively parallel capabilities of two systems[4] (McAfee, 2012). Figure 1 illustrates an architecture for Big Data analytics. Structured data are captured through various data sources including OLTP systems, legacy systems and external systems. It goes through the ETL process from the source systems to the target data warehouse. Traditional business intelligence (BI) batched analytical processing tools such as online analytical processing (OLAP), data mining, and query and reporting, can be used to create the business intelligence to enhance business operations and decision processes. Unstructured and semi-structured Big Data sources can be of a wide variety that includes data from clickstreams, social media, machine-to-machine, mobile device, sensors, documents and reports, Web logs, call records, scientific research, satellites, and geospatial devices. They are loaded into the Hadoop Distributed File System cluster.

Hadoop MapReduce provides the fault-tolerant distributed processing framework across the Hadoop cluster, where batched analytics can be performed. Actionable insight resulting from Hadoop MapReduce analytics and business intelligence analytics can be consumed by operational and analytical applications.

While Hadoop is highly scalable and can perform massively parallel computing for Big Data, it is a batch system with high latency, and would not be suitable for processing of real-time events. [3] (Chan, 2013) described geospatial intelligence as using data about space and time to improve the quality of predictive analysis. For example, real-time recommendations of places of interest can be based on the real-time location from smartphone usage. This real-time information can be combined with batched analytics to improve the quality of the predictions. Other examples of real-time analytic applications include real-time trending of social media data, real-time Web click stream analysis, algorithmic trading, and real-time M2M analysis. Emerging technologies for Big Data real-time analytics include technologies for collection and aggregation of real-time data for Hadoop, in-memory analytic systems, and real-time analytics applications for processing of data stored in Hadoop. Real-time insight created by real-time analytics can be consumed by real-time operations and decision processes.

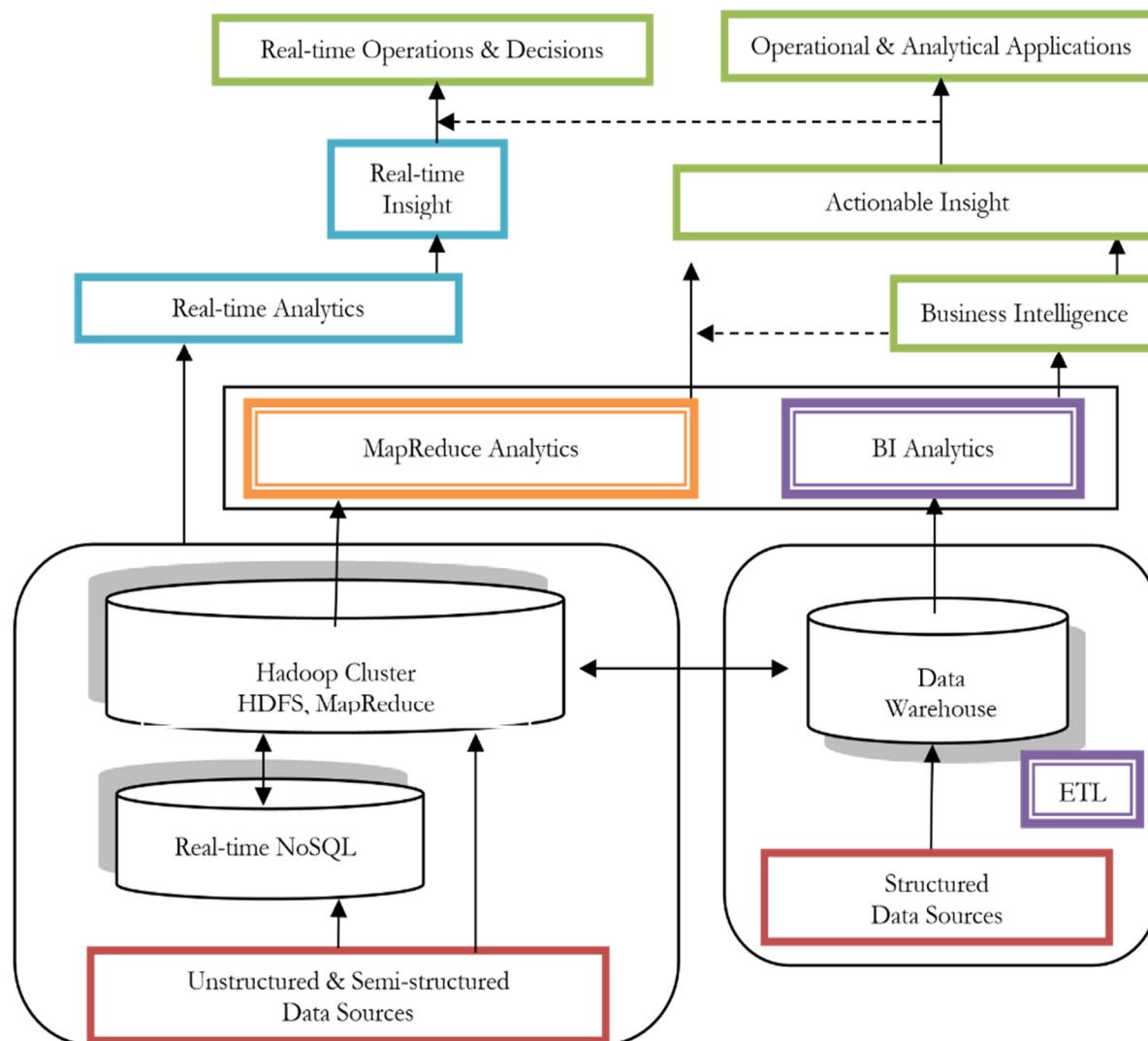


Figure 1: An Architecture for Big Data Analytics[3]

### Big Data in Agriculture

There is a number of definitions exists about big data, analyst firm Gartner defines big data as "high-volume, high velocity and high-variety information assets that demand cost-effective, innovative forms of information processing for enhanced insight and decision making." [5] Hypothesize that three characteristics, volume, velocity, and variety, make the field of big data fundamentally different from analytics and hence companies need a different set of strategies to become big data enabled. In addition to the technical challenges, big data also presents a set of managerial challenges. Most important among these is the organization's newfound ability to make data-driven decisions and hence move away from intuition-based decision making.

On the agricultural side, [6] of Business Week reports of innovative company, Climate Corp which uses localized weather data to predict crop yields and uses this data to modify crop insurance premiums to individual farmers. Climate Corp stores and analyzes more than **60** years of weather data consisting of precipitation, temperature, and soil conditions with a granularity of 2 1/2-mile radius in order to arrive at the crop yield forecasts[6].

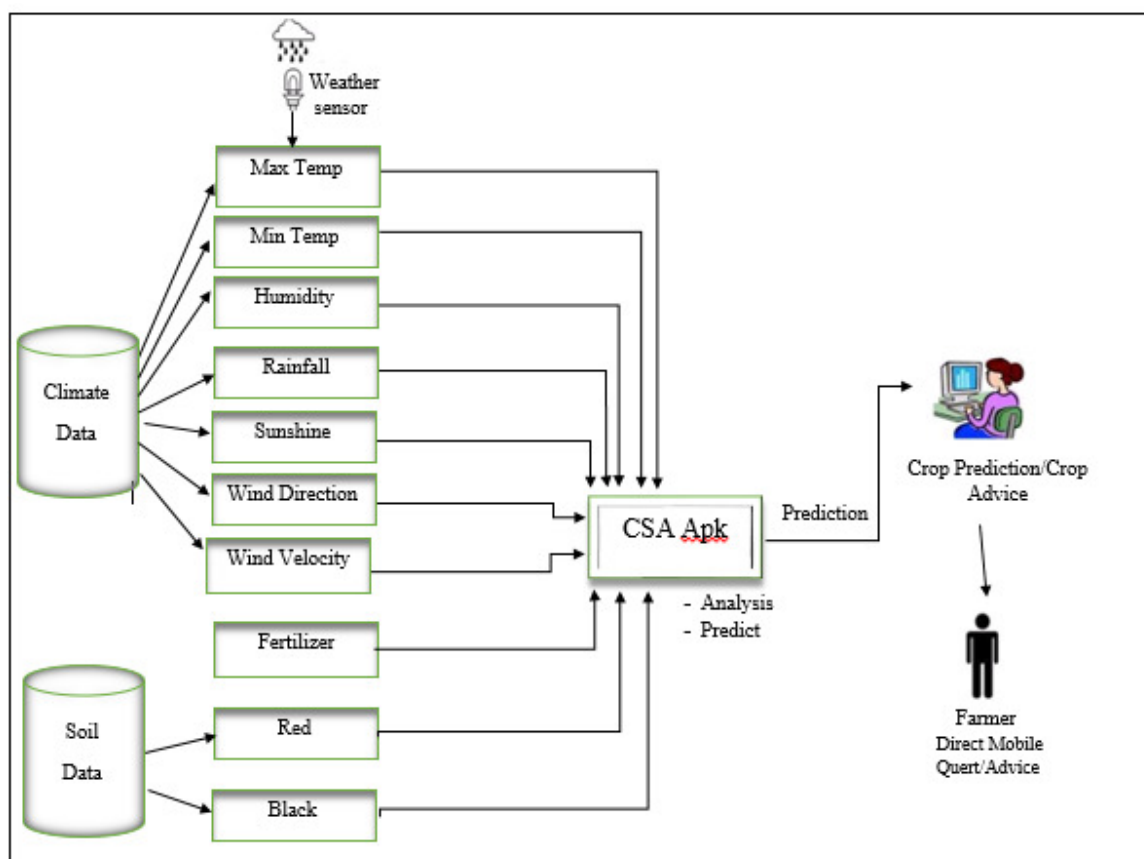


Figure 2: Framework of the CSA Application

### Machine Learning Algorithms

Currently, non-cloud machine learning libraries lack scalability, are not well tested and not built with production-quality, and they also lack open source community support. Therefore, Mahout has been created with scalable machine-learning algorithms all written using Map-Reduce paradigm, including (a) supervised learning algorithms such as neural networks, support vector machines, Naive Bayesian classifiers, decision trees, random forests, and logistic regression; (b) unsupervised learning algorithms including k-means, hierarchical clustering, self-organizing maps, fuzzy k-means, etc. Hundreds of ML algorithms have been tested by Mahout Libraries, Best ML algorithms are automatically selected by Algorithm Selection Optimizer to address specific needs of users. The selected machine learning algorithms are executed by Intelligent Mobile Agents in Map-Reduce paradigm to enable computational intelligence, real time big data analytics, and scalable computing algorithms[7] (Brian H Xu).

### Machine Learning for Big Data Analytics

Machine learning methods are very effective and relevant to many real world applications such as network security, bioinformatics, healthcare, banking and finance, and transportations. The bioinformatics and health related data are created and accumulated continuously, resulting in large volume of data. Newer forms of big data, such as 3D imaging, genomics and biometric sensor readings are also fueling this exponential growth. Future applications of real-time data, such as early detection of infections/diseases and fast application of the appropriate treatments (not just broad-spectrum antibiotics) could reduce patient morbidity and mortality. The ability to perform real-time analytics against such voluminous stream data across all specialties would help weather climate prediction.

Machine learning is a field of computer science that studies the computational methods that learn from data [8]. There are two types of learning methods in machine learning; supervised and unsupervised learning methods[9]. In supervised learning, a method learns from a set of objects with class label, often called a training set. The acquired knowledge is used to assign label to unknown objects often called test objects. Whereas unsupervised learning methods do not depend on the availability of prior knowledge or training instances with class labels. Both of methods require preprocessing of datasets for actual results. Feature selection is one of the important preprocessing tasks that leads to improved result and reduced time requirement. So Hybrid learning methods, such as Deep learning, provide significantly high accuracy in recent worlds.

## Hadoop Ecosystem & Tools

A big data platform consists of a number of components and is often called a big data ecosystem. As [10] notes, a big data platform could be compared to an oil refinery: a significant number of inputs, output, processes and structures guide the process of moving, storing, and analyzing data.

Hadoop ecosystem includes a set of tools that function near MapReduce and HDFS (the two main Hadoop core components) and help the two store and manage data, as well as perform the analytic tasks. As there is an increasing number of new technologies that encircle Hadoop, it is important to realize that certain products maybe more appropriate to fulfill certain requirements than others. Hadoop core technologies provide a fault tolerance mechanism to store large datasets. *Hadoop Distributed File System (HDFS)* is where data are stored. Data files are broken into blocks and are distributed over the servers. It is designed to run on several clusters and to be resilient to failures since it makes several copies of its data blocks. MapReduce, one the other hand, is a paradigm to process data. It was the first programming method to develop applications in Hadoop, comprising of two programs written in java: *Mappers*, to extract data from HDFS and put into maps, and *Reducers* to aggregate the results produced by the mappers.

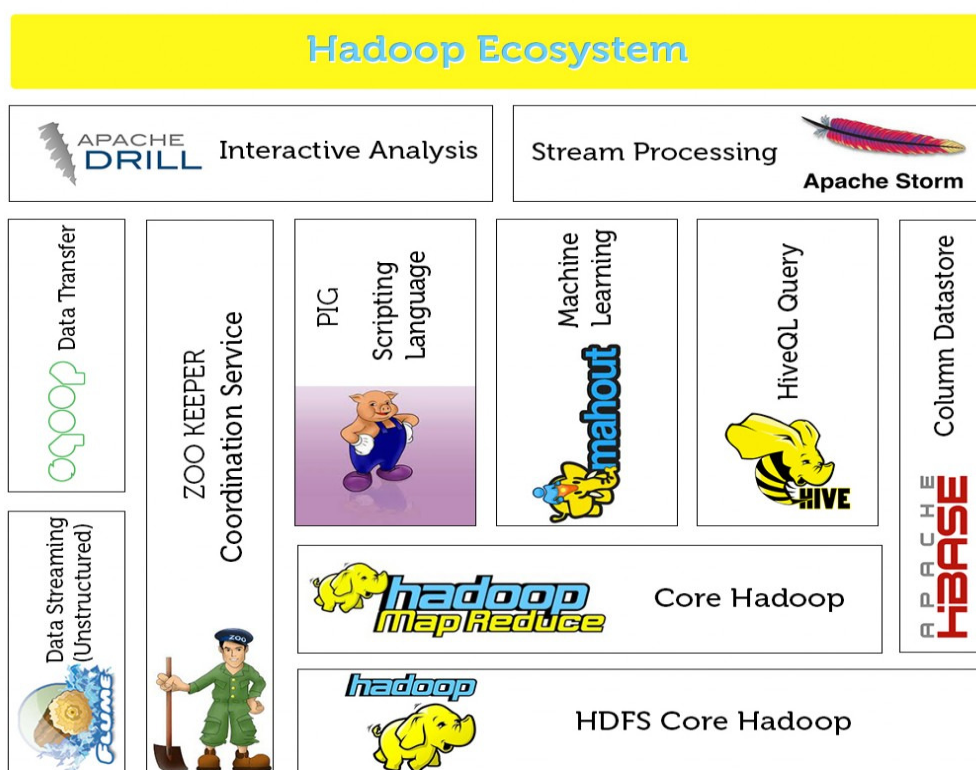


Figure 3: Hadoop Ecosystem [11]

## Conclusion

Climate-smart agriculture (CSA) is an approach for transforming and reorienting agricultural systems to support food security under the new realities of climate change. Widespread changes in rainfall and temperature patterns threaten agricultural production and increase the vulnerability of people dependent on agriculture for their livelihoods, which includes most of the world's poor. The use of Big Data Analytics and other applications to increase the agricultural yield production and reduce the expenses of the farming by taking the accurate decisions. It is important that agriculturists and the farming community in general understand the behaviour of the climatic parameters that affect crops and other plants. Knowledge of climate variability is vital in order to minimize losses and maximise production in agriculture. Medium to long-term weather predictions are useful in planning and operation of agricultural activities while the daily weather forecasts help the farmer to organise his day.

## References

- [1] F. Security, "Climate-Smart Agriculture."
- [2] S. Scherr, S. Shames, and R. Friedman, "From climate-smart agriculture to climate-smart landscapes," vol. 1, no. 1, p. 1, 2012.

- [3] J. Chan, “An Architecture for Big Data Analytics,” 2013, vol. 13, no. 2, pp. 1–14.
- [4] D. Awadallah, A., & Graham, “Hadoop and the data warehouse: When to use which. Dayton, OH: Teradata Corporation,” 2011. [Online]. Available: <http://www.teradata.com/white-papers/Hadoop-and-the-Data-Warehouse-When-to-Use-Which/>.
- [5] E. McAfee, A., & Brynjolfsson, “Big Data: The Management Revolution. Harvard Business Review,” pp. 61–68.
- [6] V. Ashlee, “Climate Corp. Updates Crop Insurance via High Tech. Business Week, Web.” [Online]. Available: <http://www.businessweek.com/articles/2012-%0A03-22/climate-corp-dot-updates-crop-insurance-via-high-tech>. [Accessed: 10-Feb-2013].
- [7] B. Xu and S. Kumar, “Big Data Analytics Framework for Improved Decision Making.”
- [8] Et and Bishop, “Pattern recognition and machine learning,” *springer New Yor*, vol. 04, no.4, 2006.
- [9] D. Bhattacharyya and J. K. Kalita, “Network anomaly detection: A machine learning perspective,” CRC Press, 2013.
- [10] H. Inc., “Apache Hadoop: the big data refinery,” 2012. [Online]. Available: <http://info.hortonworks.com/Big-Data-Refinery-White-Paper.html>.
- [11] P. Rajan, “Hadoop: Open Source Big Data Analytics Tools.” .